GET THE PROGRAM "INVT"

Ch. 23: Inference about means

Inference:

* Confidence intervals and tests of significance

* Definition: Making conclusions about a population from statistics with a known degree of confidence

conf. level

α

Ch. 19 - 22: Inference about proportions

Ch. 23 - 25: Inference about means

* We will be estimating and testing a population mean μ

* We always list standard deviation with the mean σ

* Refresher: Formula for the standard deviation:

$$\sigma = \sqrt{\frac{\Sigma(X_i - \mu)^2}{n-1}} \quad df$$

* So if we don't know the mean, we CAN'T know the std. deviation.

* So we will be estimating 2 things: mean and std. deviation (μ, σ)

* What statistics do we use to estimate mean and std. deviation?

$\bar{X}$ $\quad$ S

* What is the sampling distribution for a sample mean?

$$\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

* But we don't know μ and σ, so we will use:

$$N\left(\bar{X}, s/\sqrt{n}\right)$$

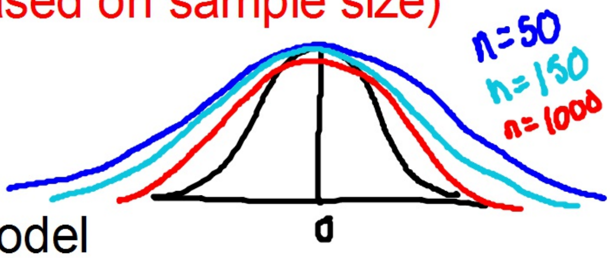
* When we do this, we cannot use Z-scores. We need a new model.

Student's $\boxed{\text{t distribution}}$

Student's t-Model:

* Family of distributions (there is more than one distribution)
* Similar to Normal model (there is only 1 distribution)
	- center at 0
	- unimodal, symmetric

* Model changes based on sample size and degrees of freedom (because S changes based on sample size)



* Degrees of freedom (df) = n - 1

* Generally wider than Normal model
(2 statistics => more variation) and (S varies in each sample. σ does not!)

* The larger the sample size (df), the closer the Student's *t*-model approaches the Normal model. $df = \infty \quad t = Z$

1-sample t-interval

CONDITIONS:

1) Randomization (SRS)
2) pop $\geq 10n$
3) Normal population or $n \geq 30$

Checking #3: * normal pop stated

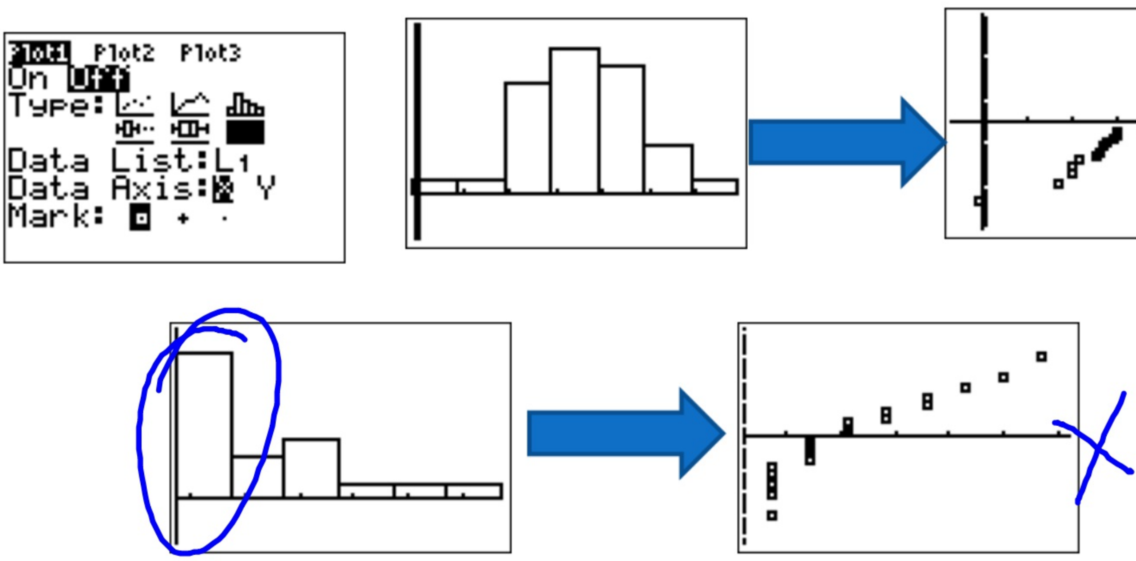
* Look at a histogram of the sample values
    * want normal shape (symmetric and unimodal)
* Look at a normal probability plot (on calculator)
    * want straight line
* Sample size greater than 30 => CLT kicks in and says sampling distribution will be normal (distribution)

** Conditions met --> Can use Student's t-model --> 1 sample t-Test/Interval

Normal Probability Plot:

* Finds the z-score of each value compared to the sample mean and std. deviation. Plots these values against themselves.

* No curve (straight line) in the plot shows a unimodal and symmetric distribution



MECHANICS:

* Calculate (or given): $\overline{x}$, s, n

write down every time

* Find degrees of freedom: df = n-1
* Find critical value t^* (A-76 or invT or Table B) form. sheet

* Formula for interval:

statistic $\pm$ (critical value)(std. dev of statistic)

$$\overline{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right) = (a, b)$$

**T-Interval on Calc.

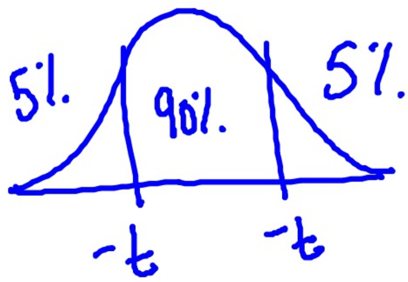
* Interpret! We are ___% confident that the true mean of ______ is between ____ and _______ units.

* invT program (not in notes)

on calculator:

2nd DISTR --> #4:invT --> ENTER

invT(% below, df) --> ENTER



program from me:

PRGM --> INVT --> ENTER

prgmINVT --> ENTER

N=? *(enter in the sample size)* 100

C=? *(enter in the confidence level)* 95%

ENTER

t = 1.984

Example: A coffee vending machine dispenses coffee into a paper cup. You're supposed to get 10 ounces of coffee, but the amount varies slightly from cup to cup. Below are the amounts measured in a random sample of 20 cups. Is there evidence that the machine is shortchanging customers? Construct a 95% confidence interval.

9.9	9.7	10.0	10.1	9.9
9.6	9.8	9.8	10.0	9.5
9.7	10.1	9.9	9.6	10.2
9.8	10.0	9.9	9.5	9.9

Conditions

1) SRS — 1) stated

2) Pop $\geq 10n$ — 2) machine makes more than 200 cups.

3) normal pop or $n \geq 30$ — 3) normal prob. plot is linear → normal pop. assumed

Conditions:

Cond. met → t distribution → 1 samp. t Interval

$$(9.845) \pm (2.093)\left(0.1986/\sqrt{20}\right)$$

$$= (9.752, 9.938)$$

We are 95% conf. that the true average amount of coffee from the machine is btw. 9.752 oz and 9.938 oz.

Mechanics & Interpretation:

1-sample t-Test

* Inference about the mean of a population (μ)

HYPOTHESES:

Ho: $\mu = \#$

Ha: $\mu > < \neq \#$

CONDITIONS:

* same as interval

(don't forget your statement afterwards!)

1 sample t-test

MECHANICS:

* Calculate (or given): $\bar{x}$, s, n

* Find: df = n - 1

* Formula:

$$t = \frac{\text{statistic - parameter}}{\text{std. dev. of statistic (SE)}} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

* P-Value: P(t $\gtrless$ test statistic) = tcdf(LB, UB, df)

$t = 3.257$

CONCLUSION:

* Compare p-value to alpha
* Same 2 sentences as always
* State conclusion in context

    - We reject/fail to reject Ho...
    - We have sufficient/insufficient evidence that the true mean of _____ is ... (use Ha)

Calculator: #2:T-Test

Example: The EPA wants to show that "the mean carbon monoxide level of air pollution is higher than 4.9." Does a random sample of 50 readings (with sample mean of 5.1 and sample std. deviation of 1.17) present sufficient evidence at the .05 level of significance to support the EPA's claim? Previous studies have indicated that such readings have an approximately normal distribution.

Cond. met → t distrib. → 1 samp t test

HYPOTHESES:

$H_0: \mu = 4.9$

$H_a: \mu > 4.9$

df = 49

n = 50

$\bar{x} = 5.1$

s = 1.17

CONDITIONS:

MECHANICS:

$$t = \frac{5.1 - 4.9}{1.17/\sqrt{50}} = 1.209$$

$$P(t > 1.209) = 0.116$$

- We fail to reject Ho....
- We have insuff. evid. that the true mean CO level is greater than 4.9 units.

CONCLUSION:

Complete practice problems #1 & 2:

1. A survey was conducted involving 250 families living in a city. The average amount of income tax paid per family in the sample was $3540 with a standard deviation of $1150. Establish and interpret a 99% confidence interval estimate for the taxes paid by families in this city.

2. The estimated U.S. intake of trans-fatty acids is 8 g per day. Consider a research project involving 150 individuals in which their daily intake of trans-fatty acids was measured. Suppose the average fatty acid intake from this sample was 12.5 g, with a standard deviation of 7.7 g. Test the research hypothesis that the average intake has increased at $\alpha = 0.05$.

1) $n = 250$ $\bar{x} = 3540$ $df = 249$ $s = 1150$ $C = 99\%$

<u>Conditions:</u>

- SRS	- assumed random
- pop $\geq$ 10n	- there are more than 2500 families in the city
- normal pop or $n \geq 30$	- $n = 250 \geq 30$

conditions met --> Student's t-distribution --> 1 samp t Int

= (3351.20, 3728.8)

We are 99% confident that the average amount of taxes paid by a family in this city is between \$3351.20 and \$3728.80.

2) $\mu = 8$ $n = 150$ $df = 149$ $\bar{x} = 12.5$ $s = 7.7$ $\alpha = 0.05$

Ho: $\mu = 8$

Ha: $\mu > 8$

Conditions:

- SRS	- assumed random
- pop $\geq$ 10n	- there are more than 1500 U.S. individuals
- normal pop or n $\geq$ 30	- n = 150 $\geq$ 30

conditions met --> Student's t-distribution --> 1 samp t Test

$$Z = \frac{12.5 - 8}{\frac{7.7}{\sqrt{150}}} = 7.158$$

$P(t > 7.158 | df = 149) = 1.737 \times 10^{-11}$

- We reject Ho b/c p-value of 1.737×10^{-11} is $< \alpha = 0.05$.
- We have sufficient evidence that the true average intake of trans-fatty acids is greater than 8 grams per day.

HW: p. 554 #2, 4, 14, 37